

Estimation of covariance components between one continuous and one binary trait

EM

H. Simianer¹ and L.R. Schaeffer²

¹ *Justus Liebig University, Department of Animal Breeding and Genetics, Bismarckstr. 16 D-6300 Giessen, FRG*

² *University of Guelph, Department of Animal and Poultry Science, Guelph, Ontario N1G 2W1 Canada*

(received 17 December 1987, accepted 14 October 1988)

Summary – A method is described to estimate variance and covariance components in a multiple trait situation with one continuous and one binary trait. An underlying bivariate normal distribution is assumed with one variable dichotomized on the observable scale through a fixed threshold. A mixed linear model is applied to the underlying scale, and Bayesian arguments are employed to derive estimation procedures for both location and dispersion parameters. This leads to a nonlinear system of equations similar to the mixed model equations for observations that have been transformed by a Cholesky decomposition of the residual variance-covariance matrix so that the residual covariance between the two transformed traits is zero, thereby simplifying construction of the multiple trait mixed model equations. The procedures for estimating genetic variances and covariances and the residual variance for the continuous trait are equivalent to restricted maximum likelihood in the multivariate normal case. The residual correlation is estimated using a maximum likelihood approach. Suitable computing strategies are indicated and a simulation study is given to illustrate the use of the method. The impact of small subclass size on the estimates is seen to be a serious drawback to the proposed method. Possible generalizations of the method and potential problems in its practical application are discussed.

variance – covariance – estimation – binary trait methods

Résumé – Estimation des composantes de la covariance entre un caractère continu et un caractère binaire. On décrit une méthode pour estimer les composantes de variance et de covariance, dans le cas de deux caractères, l'un continu et l'autre binaire. On suppose l'existence d'une distribution sous-jacente binormale, où l'une des variables ne présente que deux états observables en fonction de sa valeur par rapport à un seuil fixe. On écrit un modèle linéaire mixte pour les variables continues sous-jacentes, et l'on s'appuie sur une approche bayésienne pour construire des estimateurs des paramètres de position et de dispersion. On obtient un système d'équations non linéaires semblable aux équations d'un modèle mixte; une simplification des équations est obtenue si la covariance résiduelle entre les deux caractères est nulle, ce qui s'obtient en réalisant au préalable une décomposition de Cholesky de la matrice des variances et covariances résiduelles. La procédure d'estimation des variances et covariances génétiques, et de la variance résiduelle du caractère continu est équivalente à celle d'un maximum de vraisemblance restreint dans le cas normal; la

corrélation résiduelle est estimée selon le principe du maximum de vraisemblance. On indique des stratégies adaptées au calcul, et une simulation illustre l'utilisation de la méthode. Cette méthode s'avère sensible à l'existence de petits nombres d'observations dans certaines cellules. On discute les généralisations possibles de la méthode, ainsi que les problèmes potentiels de son application pratique.

composantes de la variance – composantes de la covariance – estimation – caractère binaire

INTRODUCTION

Discrete traits are predominant in certain areas of animal production, *e.g.*, fertility, prolificacy, viability and animal health. These traits are of increasing importance since many producers are under quota systems and a means of increasing income is to improve efficiency through these non-production traits. However, traditional traits like milk yield, milk composition, growth and feed efficiency are still important, and so are the relationships between the discrete and the traditional traits, most of which are continuous. Non-linear procedures based on the threshold concept (Dempster & Lerner, 1950) and Bayesian arguments have become available to analyze categorical observations (Gianola & Foulley, 1983; Harville & Mee, 1984). Bayesian methodology has been established as a general framework for the analysis of any type of observation arising in animal production (Foulley & Gianola, 1984; Gianola *et al.*, 1986; Höschele *et al.*, 1986, 1987; Foulley *et al.*, 1987a, 1987b).

Observation structures that include both continuous and binary traits have been discussed as far as the estimation of genetic and environmental effects is concerned (Foulley *et al.*, 1983). The objective of this paper is to present a method for the estimation of dispersion parameters of the joint distribution of continuous and binary traits. A simulation study was conducted to study the effects of small subclass size on parameter estimates.

MODEL AND DATA STRUCTURE

Each animal is recorded for two different traits. Let y_{i1} be the continuous trait 1 (*e.g.*, milk yield) for the i -th animal, and let c_i be the response for binary trait 2 (*e.g.*, mastitis), which is either 1 or 0. Categories of response for the binary trait are assumed to be mutually exclusive and exhaustive. Observations for trait 1 form the $s \times 1$ vector \mathbf{y}_1 , and the observations for the binary trait can be represented as an $s \times 1$ vector, $\mathbf{c}' = (c_1, c_2, \dots, c_s)'$, where $c_i = 0$ if no disease is observed, otherwise $c_i = 1$.

A non-observable underlying continuous variate, y_{i2} , is assumed, and y_{i1} and y_{i2} follow a multivariate normal distribution. According to the threshold concept (Dempster & Lerner, 1950), y_{i2} can be thought of as liability to contract the disease. Only if the value of y_{i2} exceeds a fixed threshold on the underlying scale, does animal i show the disease.

Let \mathbf{y}_2 be an $s \times 1$ vector containing y_{i2} for animals 1 to s , then the linear bivariate model in matrix notation is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \quad (1)$$

where :

- $\boldsymbol{\beta}_j = p \times 1$ vector of fixed effects for trait j ,
- $\mathbf{u}_j = q \times 1$ vector of additive genetic effects for trait j ,
- $\mathbf{e}_j = s \times 1$ vector of residuals for trait j ,
- $\mathbf{X} = s \times p$ incidence matrix corresponding to $\boldsymbol{\beta}_j$,
- $\mathbf{Z} = s \times q$ incidence matrix corresponding to \mathbf{u}_j ,
- $p =$ the number of levels of all fixed effects, and
- $q =$ the number of additive genetic effects, which could be greater than or equal to s .

The expectations and variance-covariance matrices of the random variables are:

$$\begin{aligned} E \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{X}\boldsymbol{\beta}_1 \\ \mathbf{X}\boldsymbol{\beta}_2 \end{bmatrix}, E \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, E \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \\ V \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} &= \mathbf{G} * \mathbf{A}, \mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}, \\ V \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} &= \mathbf{R} * \mathbf{I}, \mathbf{R} = \begin{bmatrix} r_{11} & \rho_e(r_{11}r_{22})^{1/2} \\ \rho_e(r_{11}r_{22})^{1/2} & r_{22} \end{bmatrix}, \end{aligned}$$

where:

- $\mathbf{G} = 2 \times 2$ additive genetic variance-covariance matrix,
- $\mathbf{A} = q \times q$ additive genetic relationship matrix,
- $\mathbf{R} = 2 \times 2$ residual variance-covariance matrix,
- $\mathbf{I} = s \times s$ identity matrix,
- $*$ = direct (Kronecker-) product,

- $g_{jj} =$ additive genetic variance of trait j ,
- $r_{jj} =$ residual variance of trait j ,
- $g_{12} =$ additive genetic covariance between traits 1 and 2, and
- $\rho_e =$ residual correlation between traits 1 and 2.

The unknown parameters are the location parameters,

$$\boldsymbol{\theta}' = \{\boldsymbol{\beta}', \mathbf{u}'\}; \boldsymbol{\beta}' = \{\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2\}; \mathbf{u}' = \{\mathbf{u}'_1, \mathbf{u}'_2\},$$

and the dispersion parameters

$$\boldsymbol{\gamma}' = \{\mathbf{g}', \mathbf{r}'\}; \mathbf{g}' = \{g_{11}, g_{12}, g_{22}\}; \mathbf{r}' = \{r_{11}, \rho_e\}.$$

The residual variance structure is a function of only two parameters, because y_{i2} is a conceptual variable and, therefore, an arbitrary value for r_{22} can be chosen.

Note that this model assumes that every animal is observed for both traits and that the same model applies to each trait.

METHODS OF INFERENCE

Assume that $\boldsymbol{\beta}$ has a flat prior distribution (*i.e.*, nothing is known *a priori* about the elements of $\boldsymbol{\beta}$), and that \mathbf{u} has a multivariate normal distribution with expectation null and variance-covariance matrix equal to $\mathbf{G} * \mathbf{A}$. The vectors $\boldsymbol{\beta}$ and \mathbf{u} are a

priori assumed to be independent. The joint posterior distribution of all unknowns can be written as:

$$f(\boldsymbol{\theta}, \boldsymbol{\gamma} \mid \mathbf{Y}, y_1) \sim f(\mathbf{Y}, y_1 \mid \boldsymbol{\theta}, \mathbf{r}) \cdot f(\mathbf{u} \mid \mathbf{g}) \cdot f(\boldsymbol{\gamma}) \tag{2}$$

where $f(\boldsymbol{\gamma})$ is the joint prior density of the dispersion parameters (Foulley *et al.*, 1987a).

Estimation of location parameters

Integrating $\boldsymbol{\gamma}$ out of (2) would hardly be possible, as was pointed out by Gianola *et al.* (1986). An alternative is to replace $\boldsymbol{\gamma}$ by some value, $\boldsymbol{\gamma}^{[t]}$, so that

$$f(\boldsymbol{\theta} \mid \mathbf{Y}, y_1, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}) \sim f(\mathbf{Y}, y_1 \mid \boldsymbol{\theta}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}) \cdot f(\mathbf{u} \mid \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}) \tag{3}$$

as elaborated by Foulley *et al.* (1983). Ignoring the superscript $[t]$ and assuming that there is only one continuous and one binary trait and that both traits are described by the same model, then following the arguments of Foulley *et al.* (1983), the residual variance of the conceptual variable is chosen to be:

$$r_{22} = (1 - \rho_e^2)^{-1}$$

and therefore:

$$\mathbf{R} = \begin{bmatrix} r_{11} & \rho_e(r_{11})^{1/2}(1 - \rho_e^2)^{-1/2} \\ \rho_e(r_{11})^{1/2}(1 - \rho_e^2)^{-1/2} & (1 - \rho_e^2)^{-1} \end{bmatrix}.$$

For $|\rho_e| < 1$, \mathbf{R} is positive definite and a lower triangular matrix \mathbf{T} exists, so that $\mathbf{TT}' = \mathbf{R}$. \mathbf{T}^{-1} is used to perform a Cholesky transformation to remove the residual covariance between transformed variables where

$$\mathbf{T} = \begin{bmatrix} (r_{11})^{1/2} & 0 \\ \rho_e(1 - \rho_e^2)^{-1/2} & 1 \end{bmatrix} \text{ and}$$

$$\mathbf{T}^{-1} = \begin{bmatrix} (r_{11})^{-1/2} & 0 \\ -\rho_e(r_{11})^{-1/2}(1 - \rho_e^2)^{-1/2} & 1 \end{bmatrix}.$$

Let the tilde symbol, \sim , above a variable indicate that same variable on the transformed scale, then

$$\begin{bmatrix} \tilde{y}_{i1} \\ \tilde{y}_{i2} \end{bmatrix} = \mathbf{T}^{-1} \begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix}$$

for $i = 1$ to s , and similarly for elements of $\boldsymbol{\beta}$, \mathbf{u} , and \mathbf{e} . On the transformed scale, variances and covariances are

$$\tilde{\mathbf{G}} = \mathbf{T}^{-1} \mathbf{G} \mathbf{T}'^{-1} \text{ and}$$

$$\tilde{\mathbf{R}} = \mathbf{T}^{-1} \mathbf{R} \mathbf{T}'^{-1} = \mathbf{T}^{-1} \mathbf{T} \mathbf{T}' \mathbf{T}'^{-1} = \mathbf{I}.$$

Using $b = \rho_e(r_{11})^{-1/2}(1 - \rho_e^2)^{-1/2}$

Foulley *et al.* (1983) show that under multivariate normality

$$E(y_{i2} \mid \boldsymbol{\theta}, y_{i1}) = \mu_i = \mathbf{x}'_i \tilde{\boldsymbol{\beta}}_2 + \mathbf{z}'_i \tilde{\mathbf{u}}_2 + by_{i1} \tag{5}$$

where \mathbf{x}'_i and \mathbf{z}'_i are rows of \mathbf{X} and \mathbf{Z} pertaining to animal i ,

$$\begin{aligned} (\tilde{\beta}_1 \tilde{\beta}_2)' &= \mathbf{T}^{-1}(\beta_1 \beta_2)' \text{ and} \\ (\tilde{u}_1 \tilde{u}_2)' &= \mathbf{T}^{-1}(u_1 u_2)'. \end{aligned}$$

As

$$V(e_{i2} | e_{i2}) = 1,$$

and the residuals on the transformed scale are uncorrelated, one can proceed as in a single binary trait analysis, a simplification of (Gianola & Foulley, 1983) from several categories to two, computing

$$\begin{aligned} v_i &= \phi(\mu_i)/\Phi(\mu_i) \text{ if } c_i = 1, \text{ or} \\ v_i &= -\phi(\mu_i)/(1 - \Phi(\mu_i)) \text{ if } c_i = 0, \text{ and} \\ w_i &= v_i^2 + \mu_i v_i \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution function of a standard normal variate, respectively.

Let:

$$E(y_{i2} | \theta, y_{i1}, c_i) = x'_i \tilde{\beta}_2 + z'_i \tilde{u}_2 + by_{i1} + v_i/w_i$$

then from (4),

$$\tilde{y}_{i2} = -by_{i1} + y_{i2}, \text{ and}$$

$$\begin{aligned} E(\tilde{y}_{i2} | \theta, y_{i1}, c_i) &= -by_{i1} + E(y_{i2} | \theta, y_{i1}, c_i) \\ &= x'_i \tilde{\beta}_2 + x'_i \tilde{u}_2 + v_i/w_i. \end{aligned}$$

The mode of the posterior density (3) can be calculated by applying the Newton-Raphson algorithm to the log of (3), which leads to the nonlinear system of equations for the $[k + 1]$ th round of iteration (Foulley *et al.*, 1983)

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & 0 & 0 \\ & \mathbf{Z}'\mathbf{Z} + \mathbf{V}^{11} & 0 & \mathbf{V}^{12} \\ & & \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} \\ \text{sym.} & & & \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{V}^{22} \end{bmatrix}^{[k]} \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{u}_1 \\ \tilde{\beta}_2 \\ \tilde{u}_2 \end{bmatrix}^{[k+1]} = \begin{bmatrix} \mathbf{X}'\tilde{y}_1 \\ \mathbf{Z}'\tilde{y}_1 \\ \mathbf{X}'\mathbf{W}\tilde{y}_2 \\ \mathbf{Z}'\mathbf{W}\tilde{y}_2 \end{bmatrix}^{[k]} \quad (6)$$

where

$\mathbf{W} = \text{diag} \{w_i\}$, an $s \times s$ diagonal matrix with w_i in the diagonal and

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}^{11} & \mathbf{V}^{12} \\ (\mathbf{V}^{12})' & \mathbf{V}^{22} \end{bmatrix} = (\tilde{\mathbf{G}} * \mathbf{A})^{-1}$$

Note, that \mathbf{W} and y_2 in round $[k + 1]$ are computed based on the solutions from the preceding round $[k]$.

Let

$$\tilde{\theta}^{[k+1]} = \{\tilde{\beta}'_1, \tilde{\beta}'_2, \tilde{u}'_1, \tilde{u}'_2, \} = \{\theta_i\}$$

and :

$$d = \{d_i\} \text{ with } d_i = (\theta_i^{[k+1]} - \theta_i^{[k]})/\theta_i^{[k]}.$$

Iteration on (6) is continued until $d' d < \varepsilon$, where ε is an arbitrarily small number. Suppose this criterion is met in round $[k']$, then the converged solutions have to

be transformed back to the original scale (the superscript $[t]$ indicating the set of dispersion parameters used):

$$\begin{bmatrix} \beta'_1 \\ \beta'_2 \end{bmatrix}^{[t]} = \mathbf{T} \begin{bmatrix} \tilde{\beta}'_1 \\ \tilde{\beta}'_2 \end{bmatrix}^{[k']} ; \begin{bmatrix} u'_1 \\ u'_2 \end{bmatrix}^{[t]} = \mathbf{T} \begin{bmatrix} \tilde{u}'_1 \\ \tilde{u}'_2 \end{bmatrix}^{[k']} .$$

Estimation of genetic variances and covariances

Foulley *et al.* (1983) discuss briefly the case of the variance-covariance structure being unknown and suggest the application of the restricted maximum likelihood (REML) algorithm described by Schaeffer *et al.* (1978) to estimate the entire genetic variance-covariance structure as well as the residual variance of the continuous trait only, and the residual covariance between continuous and binary traits.

Methods to estimate dispersion parameters have been developed from a Bayesian background for different data structures. Procedures for single polychotomous traits have been described by Harville & Mee (1984) and by Höschele *et al.* (1987). Methods for multivariate binary traits have been suggested by Foulley *et al.* (1987a), and for multivariate continuous traits by Gianola *et al.* (1986). Essentially, all authors recommend an algorithm analogous to the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) for REML and its multitrait extension, respectively.

Let

$$P(\boldsymbol{\gamma}, \mathbf{Y}, \mathbf{y}_1) = \frac{\delta}{\delta_g} \ln f(\boldsymbol{\gamma} | \mathbf{Y}, \mathbf{y}_1), \quad (7)$$

which requires integration of (2) with respect to $\boldsymbol{\theta}$. Foulley *et al.* (1987a) show that (7) can be written as

$$P(\boldsymbol{\gamma}, \mathbf{Y}, \mathbf{y}_1) = E_c \left\{ \frac{\delta}{\delta_g} \ln f(\mathbf{u} | \mathbf{g}) \right\} + \frac{\delta}{\delta_g} \ln f(\mathbf{g}) \quad (8)$$

where E_c is the expectation taken with respect to $f(\mathbf{u} | \mathbf{Y}, \mathbf{y}_1, \boldsymbol{\gamma})$. Furthermore, Foulley *et al.* (1987a) show that whenever a flat prior for \mathbf{g} is used, the joint posterior distribution of $\boldsymbol{\gamma}$ can be maximized with respect to \mathbf{g} by maximizing $E_c\{\ln f(\mathbf{u} | \mathbf{g})\}$ at each iterate. The resulting iterative scheme for the $[t + 1]$ th round is

$$g_{ii'}^{[t+1]} = [\tilde{u}'_i \mathbf{A}^{-1} \tilde{u}'_{i'} + \text{tr}(\mathbf{A}^{-1} \mathbf{C}_{ii'})]/q; \quad i, i' = 1, 2 \quad (9)$$

where $\text{tr}(\cdot)$ is the trace operator and

$$\begin{aligned} \hat{\mathbf{u}}_i &= E(\mathbf{u}_i | \mathbf{Y}, \mathbf{y}_1, \boldsymbol{\gamma}^{[t]}) \\ \mathbf{C}_{ii'} &= \text{cov}(\mathbf{u}_i, \mathbf{u}_{i'} | \mathbf{Y}, \mathbf{y}_1, \boldsymbol{\gamma}^{[t]}). \end{aligned}$$

These terms cannot be derived explicitly, but approximations have been suggested (Harville & Mee, 1984; Stiratelli *et al.*, 1984). Making use of these approximations and performing one estimation step on the transformed scale, we replace in (9)

$$g_{ii'}^{[t+1]} \text{ by } \tilde{g}_{ii'}^{[t+1]}$$

$$\hat{\mathbf{u}}_i \text{ by } \tilde{\mathbf{u}}_i^{[k']}$$

and C_{ii} , by the $\tilde{u}_i \times \tilde{u}_i$, part of the inverted left hand side of (6). The new estimates are on the transformed scale and need to be transformed back:

$$\mathbf{G}^{[t+1]} = \mathbf{T} \begin{bmatrix} \tilde{g}_{11} & \tilde{g}_{12} \\ \tilde{g}_{12} & \tilde{g}_{22} \end{bmatrix}^{[t+1]} \mathbf{T}'$$

Foulley *et al.* (1987a) have shown that if the initial \mathbf{G} is positive definite, then the subsequent estimates of \mathbf{G} are also positive definite, which holds in combination with the applied Cholesky transformation.

Estimation of the residual variance of the continuous trait

Maximum *a posteriori* predictors tend to converge to trivial results when flat priors for the dispersion parameters are used (Lindley & Smith, 1972). As an alternative Gianola *et al.* (1986) suggest estimators arising from an approximate integration of the variances. For \mathbf{g} these are equivalent to the algorithms described in section B as long as flat priors for the variances are used. For the residual variance of the continuous trait estimator is

$$\tilde{r}_{11}^{[t+1]} = (\tilde{\mathbf{e}}_1^{[t]} \tilde{\mathbf{e}}_1^{[t]} + p + q - tr(\mathbf{A}^{-1} \mathbf{C}_{11}^{[t]}) \tilde{\sigma}_u^{-2})/s.$$

which can be shown to lead to the same results as iterating on

$$\tilde{r}_{11}^{[t+1]} = (\tilde{\mathbf{y}}_1' \tilde{\mathbf{y}}_1 - \tilde{\boldsymbol{\beta}}_1^{[k']} \mathbf{X}' \tilde{\mathbf{y}}_1 - \tilde{\mathbf{u}}_1^{[k']} \mathbf{Z}' \tilde{\mathbf{y}}_1)/(s - p) \tag{10}$$

assuming full column rank of \mathbf{X} . Estimators on the original scale are obtained by:

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix}^{[t+1]} = \mathbf{T}^{[t]} \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} \\ \tilde{r}_{12} & \tilde{r}_{22} \end{bmatrix}^{[t+1]} \mathbf{T}'^{[t]}$$

which yields

$$r_{11}^{[t+1]} = \tilde{r}_{11}^{[t+1]} r_{11}^{[t]}.$$

Estimation of the residual correlation between traits

Foulley *et al.* (1983) suggest the use of restricted maximum likelihood estimation as described by Schaeffer *et al.* (1978) for multivariate normal data. However, this requires that the residual correlations between continuous and binary traits be known. Foulley *et al.* (1983) show that proportionately except for a constant.

$$\ln f(\boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{y}_1, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}) = \sum_i [(1 - c_i) \ln (\Phi(\mu_i)) + c_i \ln (1 - \Phi(\mu_i))] - 0.5 (\tilde{\mathbf{e}}_1^{[t]} \tilde{\mathbf{e}}_1^{[t]} + \tilde{\mathbf{u}}^{[t]} \mathbf{v}^{[t]} \tilde{\mathbf{u}}^{[t]}) \tag{11}$$

The problem is to find $\boldsymbol{\gamma}^{[t]}$ and $\boldsymbol{\theta}$ (which is a function of $\boldsymbol{\gamma}^{[t]}$) which maximize this log posterior density. Estimates for $\boldsymbol{\theta}$, \mathbf{g} , and r_{11} can be obtained under the pretense that $\rho_e = \rho_e^{[t]}$ similar to the one dimensional grid search technique proposed by Smith & Graser (1986). In this case the strategy would be:

- 1) choose a set of possible values of ρ_e and a set of starting values for \mathbf{g} , and r_{11} ;

- 2) for i th value of ρ_e in the set, iterate on (6) using the dispersion parameters (\mathbf{g} , r_{11} , ρ_{ei}) until convergence of $\boldsymbol{\theta}$, then do one step of (9) and (10) and continue to iterate until convergence of (\mathbf{g} and r_{11}) is achieved. Then compute the value of (11) defined as $h(\rho_{ei})$. Repeat for all ρ_{ei} in the set;
- 3) to find the ρ_{ei} that maximizes the log likelihood, fit a second degree polynomial through the points (ρ_{ei} , $h(\rho_{ei})$). Add the mode of this curve to the set of possible values of ρ_e and repeat the second step with this new value. Continue this iterative process until convergence of ρ_e is achieved.

This strategy is computationally very demanding as convergence is very slow. An alternative empirical approach using a maximum likelihood procedure to estimate ρ_e was suggested by equation 4.4 of Tate (1955) for the biserial correlation problem. The $[t + 1]$ th improved estimate of ρ_e is computed as:

$$\rho_e^{[t+1]} = \rho_e^{[t]} + \frac{(1 - \rho_e^{2[t]}) \tilde{\mathbf{y}}_1' \mathbf{v}^{[k]}}{\tilde{\mathbf{y}}_1' \mathbf{W}^{[k']} \tilde{\mathbf{y}}_1 - \rho_e^{[t]} \Omega^{[k']} \mathbf{I}' \mathbf{W}^{[k']} \tilde{\mathbf{y}}_1} \quad (12)$$

where

$\mathbf{v} = s \times 1$ vector with elements v_i , $i = 1, s$,

$\mathbf{I} = s \times 1$ vector of I 's,

$\tilde{\mathbf{y}}_1 = (\mathbf{y}_1 - \mathbf{I} * \bar{y}_1) * I / (r_{11}^{[t]})^{1/2}$,

$\bar{y}_1 =$ phenotypic mean of y_{i1} for all animals,

and $\Omega^{[k']}$ is the estimated value of the threshold which can be calculated from $\beta_1^{[k']}$ analogously to a least squares mean. Note that this procedure yields a new estimate of ρ_e on the original scale, and therefore, no backtransformation is needed.

Computing algorithm

Combining all the different steps described, an obvious computing strategy would be:

- 1) choose a set of starting values $\boldsymbol{\gamma}^{[1]}$, set $[t] = 1$;
- 2) iterate on (6) until convergence to get $\boldsymbol{\theta} \mid \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$;
- 3) do one step each of (9), (10) and (12) to get a new set of dispersion parameters $\boldsymbol{\gamma}^{[t+1]}$, set $[t] = [t + 1]$ and continue with the precedent step.

Iterate on the above scheme until the estimates of $\boldsymbol{\gamma}$ are converged. Empirically, (10) and (12) were found to converge more quickly than (9), and the estimates in \mathbf{r} seemed to be relatively independent of those in \mathbf{g} . This suggests the use of a different strategy:

- 1) choose a set of starting values $\mathbf{g}^{[1]}$ and $\mathbf{r}^{[1]}$, set $[t] = 1$ and $[s] = 1$;
- 2) do one round of (6) to get an improved estimate of $\boldsymbol{\theta} \mid \mathbf{g} = \mathbf{g}^{[t]}$, $\mathbf{r} = \mathbf{r}^{[s]}$;
- 3) do one step each of (10) and (12) to get a new set of parameters $\mathbf{r}^{[s+1]}$, set $[s] = [s + 1]$ and continue with the precedent stage
- 4) if convergence in the second step is reached, do one step of (9) to get a new estimate of $\mathbf{g}^{[t+1]}$, set $[t] = [t + 1]$ and continue with the second step.

Iteration is stopped when the estimates in \mathbf{g} are converged. This algorithm was found to be much faster than the first, which is paralleled by the findings of Foulley

et al. (1987a) for the multiple binary trait case. Iterating on \mathbf{r} only for some rounds with every new value of \mathbf{g} improved the procedure even further, as the estimates of \mathbf{r} converge quite rapidly.

SIMULATION STUDY

Model and methods

A simulation study was done to assess the sampling properties of the proposed estimators. The assumed true model was:

$$y_{ijkl} = \mu_i + h_{ij} + s_{ik} + e_{ijkl} \quad (13)$$

where:

- y_{ijkl} = phenotypic record ($i = 1$) or underlying liability value ($i = 2$) of animal l in herd j with sire k ,
- μ_i = mean of trait i ,
- h_{ij} = effect of herd j on trait i ,
- s_{ik} = effect of sire k on trait i , and
- e_{ijkl} = residual term.

Pairs of herd, sire and residual effects were generated for each animal as random samples from a bivariate normal distribution. Trait 2 was then dichotomized, applying a threshold at 1.282 standard deviations, which assigns 90 and 10 per cent of the observations to the phenotypic classes 0 and 1, respectively. Records were generated for 10,000 individuals per replicate, which were randomly assigned to sires (of which there were 50) and to herds (of which there were either 100 or 1 000). With only 100 herds there were 13.5% empty sire by herd subclasses, and with 1 000 herds there were 81% empty sire by herd subclasses. The objective of having two data sets was to study the effect of small sire by herd subclass size on the parameter estimates from the above methods.

The true dispersion parameters were:

$$\boldsymbol{\gamma}' = \{h_1^2, h_2^2, r_g, \rho_e, r_{11}\} = \{0.25, 0.10, 0.25, 0.375, (670.0)^2\}.$$

The model of analysis was equal to (13), except that herd effects were treated as fixed and the non observable y_2 was replaced by the phenotypic observation in the binary trait. The starting values of iteration were

$$\boldsymbol{\gamma}'^{[1]} = \{0.20, 0.15, 0.0, 0.0, (600.0)^2\},$$

and the stopping criterion was

$$\sum_{i=1}^5 \left[\frac{\gamma_i^{[t]} - \gamma_i^{[t+1]}}{\gamma_i^{[t]}} \right]^2 < 10^{-7}$$

considering the relative change to account for the different magnitude of the true values. Due to computing time restrictions, only twenty-five replications of each pair of data sets were generated and analyzed.

Results

The mean estimates and 95% confidence intervals are given in Table I. Fisher's z -transformation for correlations was applied to h_1^2 , h_2^2 , r_g and ρ_e , and a t -test was applied to examine differences between true and estimated parameters. Estimates from data set I with 100 herds agreed quite closely with the true parameters. The only significant bias was observed for the residual correlation, which is partly due to the small confidence interval for this estimate. The genetic correlation had a very large confidence interval compared with the other estimates.

The results for data set II (1 000 herds) were similar with the exception of a 50% overestimation of the heritability of the binary trait. Höschele *et al.* (1987) reported similar results in a simulation study considering estimation of heritability for binary traits. The bias was observed when the average subclass size was below 2, which agrees with the present results, as the mean subclass sizes were approximately 2.3 and 1.1 in data set I and II, respectively. Höschele *et al.* (1987) considered the approximation of the conditional distribution in the derivation of (9) as the main cause of the bias, as this approximation is based on a normality assumption (Stiratelli *et al.*, 1984), which does not hold with small cell size.

Table I. True and mean estimated dispersion parameters, confidence intervals and significance level of bias from 25 replicates of data sets I and II.

Parameter	True Value	Estimated value [95% confidence interval]	
		Data Set I (100 herds)	Data Set II (1 000 herds)
h_1^2	0.25	0.2557 [0.2362; 0.2750]	0.2584 [0.2400; 0.2767]
h_2^2	0.10	0.1038 [0.0862; 0.1213]	0.1539 *** [0.1310; 0.1767]
r_g	0.25	0.2522 [0.1368; 0.3624]	0.2353 [0.1137; 0.3515]
ρ_e	0.375	0.3539 *** [0.3457; 0.3620]	0.3839 * [0.3765; 0.3912]
$(r_{11})^{1/2}$	670.0	668.4681 [666.69; 670.245]	668.2540 [666.40; 670.10]

Significance level * = $p \leq 0.05$; *** = $p \leq 0.001$

DISCUSSION

A method has been presented for the estimation of variances and covariances for the simplest case of one continuous and one binary trait. Theoretically, generalization to more complex observation structures is straightforward. For the case of n continuous and one binary trait, the method to estimate location parameters is as given by Foulley *et al.* (1983), and the procedures to estimate dispersion parameters can be readily extended, using results of Hannan & Tate (1965) for the maximum likelihood estimation of residual correlations. Further generalizations, including the use of informative priors, may be derived through Bayesian arguments as employed by Gianola *et al.* (1986) and Foulley *et al.* (1987a).

In practice, however, an analysis of even the simplest case poses a formidable computational task with a large body of data, in which case the optimization of computing strategies is of crucial importance. Actually, the methodology presented was developed for a joint analysis of more than 200,000 records on production and disease data in dairy cattle.

Observations on both traits for all animals were assumed complete throughout the data, which seldom is the case in practice, where selection and hence missing data due to selection play an important role. This has been addressed by Henderson (1975) for continuous traits and by Foulley & Gianola (1986) for categorical traits. The multiple trait methodology allows for sequential selection where the decision whether an animal is given the opportunity to be observed for the discrete trait is made on the basis of its performance for the continuous trait, but not *vice versa*, as was indicated by Foulley *et al.* (1983). General approaches to the more complex types of selection are needed, as most of the reasons for inevitable natural selection (diseases, fertility) are of categorical nature.

The simulation study was an illustration of the methodology for two special cases. A general assessment of the sampling properties of the estimates would require a series of simulations over a wide range of parameter combinations and more realistic population structures as was done by Höschele *et al.* (1987), for the single categorical trait case. Since the general framework of the methods is the same, a similar behaviour of the estimates can be expected, which was shown for the biased estimate of heritability for the binary trait when the average size of the smallest subclass was less than two. This is a serious problem in practical applications, *e.g.* when dairy progeny test data are analyzed by a model in which sire and herd \times year \times season effects are cross-classified.

The complexity of the Bayesian approach in the context of the given model and observation structure makes it necessary to use certain assumptions and restrictions as discussed by Gianola *et al.* (1986) for the multivariate normal case. Although the approach of basing the inference about θ on the conditional distribution $f(\theta | \mathbf{Y}, \mathbf{y}_1, \boldsymbol{\gamma}^*)$ where $\boldsymbol{\gamma}^*$ is the mode of $f(\boldsymbol{\gamma} | \mathbf{Y}, \mathbf{y}_1)$ has proven to be very useful (Gianola *et al.*, 1986; Höschele *et al.*, 1987), alternative approaches are possible as discussed by Foulley *et al.* (1987a). Thus, further development of Bayesian methodology may suggest changes in the proposed algorithm.

ACKNOWLEDGEMENT

This study was carried out while H. Simianer was visiting the University of Guelph, supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

- Dempster E.R. & Lerner I.M. (1950) Heritability of threshold characters. *Genetics* 35, 212-235
- Dempster A.P., Laird N.M. & Rubin D.B. (1977) Maximum likelihood from incomplete *data via* the EM algorithm. *J.R. Statist. Soc. B* 39, 1-38
- Foulley J.L., Gianola D. & Thompson R. (1983) Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genet. Sel. Evol.* 15, 401-424
- Foulley J.L. & Gianola D. (1984) Estimation of genetic merit from bivariate "all or none" responses. *Genet. Sel. Evol.* 16, 285-306
- Foulley J.L. & Gianola D. (1986) Sire evaluation for multiple binary responses when information is missing on some traits. *J. Dairy Sci.* 69, 2681-2695
- Foulley J.L., Im S., Gianola D. & Höschele I. (1987a) Empirical Bayes estimation of parameters for n polygenic binary traits. *Genet. Sel. Evol.* 19, 197-224
- Foulley J.L., Gianola D. & Im S. (1987b) Genetic evaluation for traits distributed as Poisson-binomial with reference to reproductive traits. *Theor. Appl. Genet.* 73, 870-877
- Gianola D. & Foulley J.L. (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15, 201-224
- Gianola D., Foulley J.L. & Fernando R.L. (1986) Prediction of breeding values when variances are not known. *Genet. Sel. Evol.* 18, 485-498
- Hannan J.F. & Tate R.F. (1965) Estimation of the parameters for a multivariate normal distribution when one variable is dichotomized. *Biometrika* 52, 664-668
- Harville D.A. & Mee R.W. (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408
- Henderson C.R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423-447
- Höschele I., Foulley J.L., Colleau J.J. & Gianola D. (1986) Genetic evaluation for multiple binary responses. *Genet. Sel. Evol.* 18, 299-320
- Höschele I., Gianola D. & Foulley J.L. (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *Z. Tierz. Zuechtungsbiol.* 104, 334-349
- Lindley D.V. & Smith A.F.M. (1972) Bayes estimates for the linear model. *J.R. Statist. Soc. B*, 34, 1-41
- Schaeffer L.R., Wilton J.W. & Thompson R. (1978) Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* 34, 186-193

- Simianer H. (1986) Restricted maximum likelihood estimation of variances and covariances from selected data. *Third World Congress on Genetics Applied to Livestock Production*, Lincoln, Nebraska, 16-22 July 1986, University of Nebraska, Lincoln, 12, 421-426
- Smith S.P. & Graser H.U. (1986) Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* 69, 1156-1165
- Stiratelli R., Laird N.M. & Ware J.H. (1984) Random-effects model for serial observations with binary response. *Biometrics* 40, 961-971
- Tate R.F. (1955) The theory of correlation between two continuous variables when one is dichotomized. *Biometrika* 42, 205-216